

Assignment on Student Survey

Basak Atanu

04-25-2022

As a data science intern with newly learned knowledge in skills in statistical correlation and R programming, you will analyze the results of a survey recently given to college students. You learn that the research question being investigated is: “Is there a significant relationship between the amount of time spent reading and the time spent watching television?” You are also interested if there are other significant relationships that can be discovered? The survey data is located in this StudentSurvey.csv file.

i) Use R to calculate the covariance of the Survey variables and provide an explanation of why you would use this calculation and what the results indicate.

Table 1: Student Survey Data

TimeReading	TimeTV	Happiness	Gender
1	90	86.20	1
2	95	88.70	0
2	85	70.17	0
2	80	61.31	1
3	75	89.52	1
4	70	60.50	1
4	75	81.46	0
5	60	75.92	1
5	65	69.37	0
6	50	45.67	0
6	70	77.56	1

Correlation between TimeReading and TimeTv

```
## [1] -0.8830677
```

The correlation between TV time and Reading Time are negatively correlated, which means the students who watch more TV has less time for their study, which is true because time in a particular day is always fixed.

Correlation between Happiness and TimeTv

```
## [1] 0.636556
```

The correlation between TV time and Happiness is positively correlated.

Correlation between Happiness and TimeReading

```
## [1] -0.4348663
```

The correlation between Studytime and Happiness is negatively correlated. So happiness is less if they study more.

ii) Examine the Survey data variables. What measurement is being used for the variables? Explain what effect changing the measurement being used for the variables would have on the covariance calculation. Would this be a problem? Explain and provide a better alternative if needed.

The TimeReading is using hour where as the TimeTV variable is using minutes, Lets change this hr to minute and see the effect on correlation.

##	TimeReading	TimeTV	Happiness	Gender	TimeRD
## 1	1	90	86.20	1	60
## 2	2	95	88.70	0	120
## 3	2	85	70.17	0	120
## 4	2	80	61.31	1	120
## 5	3	75	89.52	1	180
## 6	4	70	60.50	1	240

Correlation between TimeReading and TimeTV

```
## [1] -0.8830677
```

Correlation between TimeRD and TimeTV

```
## [1] -0.8830677
```

There is no effect on covariance calculation for the change in measurements.

iii) Choose the type of correlation test to perform, explain why you chose this test, and make a prediction if the test yields a positive or negative correlation?

```
#pearson test for correlation
```

```
##  
## Pearson's product-moment correlation  
##
```

```
## data: survey$TimeRD and survey$TimeTV
## t = -5.6457, df = 9, p-value = 0.0003153
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.9694145 -0.6021920
## sample estimates:
## cor
## -0.8830677
```

##The p-value of the test is 0.0003153, which is less than the significance level $\alpha = 0.05$. We can conclude that time to read and time for TV are significantly correlated.

iv) Perform a correlation analysis of below

All variables

```
## TimeReading TimeTV Happiness Gender TimeRD
## TimeReading      1.00 -0.88 -0.43 -0.09  1.00
## TimeTV           -0.88  1.00  0.64  0.01 -0.88
## Happiness        -0.43  0.64  1.00  0.16 -0.43
## Gender           -0.09  0.01  0.16  1.00 -0.09
## TimeRD           1.00 -0.88 -0.43 -0.09  1.00
```

A single correlation between two a pair of the variables

```
## [1] -0.8830677
```

Repeat your correlation test in step 2 but set the confidence interval at 99%

```
##
## Pearson's product-moment correlation
##
## data: survey$TimeRD and survey$TimeTV
## t = -5.6457, df = 9, p-value = 0.0003153
## alternative hypothesis: true correlation is not equal to 0
## 99 percent confidence interval:
## -0.9801052 -0.4453124
## sample estimates:
## cor
## -0.8830677
```

Describe what the calculations in the correlation matrix suggest about the relationship between the variables. Be specific with your explanation.

The matrix says, there's quite a high correlation between TV time and Reading Time and the value is negative so its inversely related, i.e. higher the tv time means less reading time and vice versa.

v) Calculate the correlation coefficient and the coefficient of determination, describe what you conclude about the results.

coefficient of determination

```
## [1] 0.7798085
```

the R-Squared value ranges between 0 and 1, this means 77.9% of the variation in reading time can be explained by Television watching time.

vi) Based on your analysis can you say that watching more TV caused students to read less? Explain.

Yes, there is high negative correlation between the two which explains watching more tv causes students to read less

vii) Pick three variables and perform a partial correlation, documenting which variable you are “controlling”. Explain how this changes your interpretation and explanation of the results.

```
##   TimeReading TimeTV Happiness
## 1           1      90      86.20
## 2           2      95      88.70
## 3           2      85      70.17
## 4           2      80      61.31
## 5           3      75      89.52
## 6           4      70      60.50
```

```
## Loading required package: MASS
```

```
##
```

```
## Attaching package: 'MASS'
```

```
## The following object is masked _by_ '.GlobalEnv':
```

```
##
```

```
##      survey
```

```
## $estimate
```

```
##           TimeReading      TimeTV Happiness
```

```

## TimeReading    1.0000000 -0.8729450  0.3516355
## TimeTV         -0.8729450   1.0000000  0.5976513
## Happiness      0.3516355   0.5976513  1.0000000
##
## $p.value
##           TimeReading      TimeTV  Happiness
## TimeReading 0.0000000000 0.0009753126 0.31905895
## TimeTV      0.0009753126 0.0000000000 0.06804372
## Happiness   0.3190589526 0.0680437248 0.00000000
##
## $statistic
##           TimeReading      TimeTV  Happiness
## TimeReading  0.000000 -5.061434  1.062425
## TimeTV       -5.061434  0.000000  2.108388
## Happiness     1.062425  2.108388  0.000000
##
## $n
## [1] 11
##
## $gp
## [1] 1
##
## $method
## [1] "pearson"

```

Reading time and tv time have partial correlation of -0.8729450 with p-value 0.0009753126, so We can conclude that time to read and time for TV are having significant partial correlated when happiness is controlled